

International Journal of Modern Physics C  
 © World Scientific Publishing Company

## STATISTICAL PROPERTIES OF ONLINE AUCTIONS

ALIREZA NAMAZI

ANDREAS SCHADSCHNEIDER

*Institut für Theoretische Physik, Universität zu Köln  
 50937 Köln, Germany  
 {an,as}@thp.uni-koeln.de*

Received February 9, 2008

Revised Day Month Year

We characterize the statistical properties of a large number of online auctions run on eBay. Both stationary and dynamic properties, like distributions of prices, number of bids etc., as well as relations between these quantities are studied. The analysis of the data reveals surprisingly simple distributions and relations, typically of power-law form. Based on these findings we introduce a simple method to identify suspicious auctions that could be influenced by a form of fraud known as shill bidding. Furthermore the influence of bidding strategies is discussed. The results indicate that the observed behavior is related to a mixture of agents using a variety of strategies.

*Keywords:* Complex systems; emergent behaviour; socio-economic systems; auctions

### 1. Introduction

The Internet has triggered new forms of commerce. Not only has it become possible to order almost every product comfortably from your home, but also different forms of trading have become popular. Here especially the success of online auction sites is remarkable. Through the Internet participation in auctions is no longer just for a minority, but allows millions of consumers all over the world to buy and sell almost any kind of product. Nowadays, auction sites rank very high in the number of visitors, and also in the average time spent per visit.

Understanding the interactions between users is not only interesting from an economic point of view. Surprisingly, despite the success of econophysics<sup>1,2</sup>, only very few studies of the statistical properties of online auctions by physicists exist<sup>3,4</sup>. Ref. 3 focused on properties of agents (i.e. bidders and sellers) in auctions, e.g. the number of agents participating or the frequency of bids placed. It was found that online auctions can be considered as complex systems that exhibit emergent behaviour. In Ref. 4, the relation between different agents was investigated by analyzing the structure of the interaction network resulting from online auctions.

Online auctions are conducted in a different way than the standard English auction which usually ends when nobody is willing to bid higher ("soft close"). In

contrast, online auctions end at a fixed time known to all agents ("hard close"). Any agent offering an item at an online auction house like eBay first has to specify the starting time, starting price and duration of the auction. Each new bid placed by another agent has to exceed the currently listed price by a preset increment. Agents can either bid incrementally by placing a bid that corresponds to the current price plus the minimal acceptable increment, or take advantage of *proxy bidding*. In proxy bidding an agent indicates the maximum price he/she is willing to pay for the given item (proxy price), which is not disclosed to other bidders. Each time a bidder places a new bid, the auction house makes automatic bids for the agent with an active proxy bid <sup>a</sup>, outbidding the last bid with a fixed increment until the proxy price is reached. The agent with the highest bid wins.

Due to the proxy system eBay auctions are very similar to *second price auctions*<sup>5</sup>. The winner (buyer) does not pay a price corresponding to his actual winning bid. Instead the final price is determined by the second highest bid, plus the preset increment.

From a physics point of view the development of a simple microscopic model for online auctions appears to be quite challenging. In contrast to most other processes it is essential that the dynamics ends at a certain time and that this fact is known to all the agents. As we will show below, this is clearly reflected in the empirical data. We will not deal with the problem of modelling here, but instead try to determine the generic statistical properties of online auctions empirically. These data might then be used in order to test the validity of model approaches.

eBay keeps a detailed record of the bidding history that is publicly available. Bidding agents are distinguished by a unique user name. This allows to study the dynamics of the bidding process in a quantitative way. In this paper, we analyze and characterize the statistical properties of online auction data. We focus on the properties of the auctions and especially the relations between various quantities (price, number of bids placed etc.). Furthermore we also study the dynamics of the bidding process, not only the stationary properties after the end of the auction. This allows interesting insights into the strategies used by the agents participating in the auction.

## 2. Data collection

For our investigation we have used two major sets of auction data from eBay Germany ([www.ebay.de](http://www.ebay.de)) that allow us to focus on various properties in more detail. Many of our results are found to be valid for both data sets.

The first set (DB-1) comprises data collected from auctions running on March 22, 2004. We focus on auctions with the label "OVP" <sup>b</sup> in the title, indicating a new product. The data comprises 173,315 auctioned items, grouped in 9904 subcat-

<sup>a</sup>Note that at most one *active* proxy bid can exist at any time.

<sup>b</sup>Abbreviation for 'Originalverpackung', meaning 'original wrapping' or 'sealed'.

egories by eBay. 262,508 distinct agents bidding on items and 43,500 sellers offering auctioned items are identified.

The second set (DB-2) are data collected over 10 months in the subcategories "web projects" and "websites & domains". The auctions involve 11,145 agents bidding on 52,373 items.

The data set DB-1 allows to study auctions where the majority of products has a well-defined value (e.g. market price) for all agents. In contrast, the set DB-2 comprises items where the value can be different for different agents.

### 3. Statistical analysis

Using these data one finds that the distribution of the total number of bids placed by a given agent follows a power-law for both data sets. The total number of distinct items offered by a given agent (as seller) also follows a power-law distribution.

However, first we have examined distributions of static properties (i.e. properties at the end of an auction). A good qualitative agreement with the results of Ref. 3 is observed. E.g. the distributions of the distinct number of agents  $n_{\text{agent}}$  simultaneously bidding on a certain item and the total number of bids  $n_{\text{bids}}$  received for an item both follow exponential distributions  $P(n) \sim \exp(-n/n_0)$ , where  $n_0 = 6.5$  for  $n_{\text{bids}}$  and  $n_0 = 2.9$  for  $n_{\text{agent}}$ . This is in agreement with Ref. 3, where the values  $n_0 = 5.6$  for  $n_{\text{bids}}$  and  $n_0 = 2.5$  for  $n_{\text{agent}}$  for eBay and  $n_0 = 10.8$  for  $n_{\text{bids}}$  and  $n_0 = 7.4$  for  $n_{\text{agent}}$  for eBay Korea were obtained.

The activity of individual agents as bidder or seller follows power-law distributions<sup>6</sup>. One can find that the distribution of the total number of bids placed by the same agent, denoted by  $n_{\text{bids}}$ , follows a power law

$$P(n_{\text{bids}}) \sim n_{\text{bids}}^{-\gamma}, \quad (1)$$

where  $\gamma = 1.9$ .

In order to quantify the bidding process we define a *dimensionless* variable that we call return  $\varrho$ . It is the relative increase of the submitted bid  $b$ :

$$\varrho = \frac{b - p_{\text{current}}}{p_{\text{current}}}, \quad (2)$$

where  $p_{\text{current}}$  is the current or listed price just before the bid is placed. In analogy to the quantity used in financial markets<sup>1,2</sup> it measures relative changes. Very large values of the return are usually generated by the first submitted bid, if the bidder follows the recommendations of eBay and submits the maximum price (s)he is willing to pay. The distribution of  $\varrho$  is found to follow a power law for almost three orders of magnitude with exponent  $-2.44$  (Fig. 1). The form of the distribution appears to be quite stable even for different time spans of the bid submission<sup>6</sup>.

Although the bids are correlated, the returns show only very short-ranged correlations. By computing the correlation function  $c_{ij} = \langle \varrho_i \varrho_j \rangle - \langle \varrho_i \rangle \langle \varrho_j \rangle$  of the returns (indices  $i$  denote the chronological order of arriving bids; averaging is done over

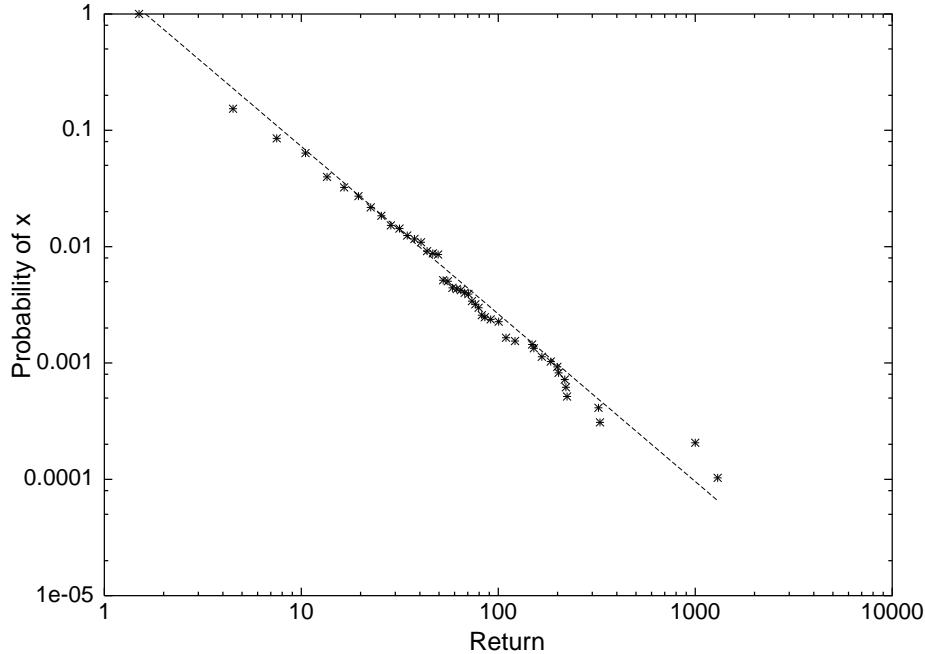


Fig. 1. The cumulative probability distribution of returns  $\varrho$  follows a power law with exponent  $-1.44$ .

all auctions) one finds that off-diagonal terms  $c_{ij}$  with  $i \neq j$  are very small<sup>6</sup>. This result is rather surprising and we have currently no simple explanation.

One of the interesting questions is the functional dependence of the price on other dynamic parameters such as number of bids. Several studies<sup>7,8,9</sup> investigate the influence of static parameters like the ending time (which day of the week, on which daytime), start price, reputation of the seller etc. Much less is known (quantitatively) about relations between dynamic parameters.

We have determined the dependence of the price on other dynamic parameters. For the relation between the final price  $p_{\text{final}}$  and number of bids  $n_{\text{bid}}$  placed on the item a power law is found:

$$\langle p_{\text{final}} \rangle \propto n_{\text{bid}}^{\alpha}, \quad (3)$$

where  $\alpha = 1.58$  for the data set DB-2 and  $\alpha = 1.53$  for the data set DB-1 (Fig. 2).

This functional dependence seems to be universal and independent of the sub-categories or time intervals used for data collection<sup>6</sup>. Especially there seems to be no significant difference between the two data sets DB-1 and DB-2.

Empirically we find that the distribution of the final price for fixed total number of bids is well described by a log-normal distribution<sup>6</sup>. A possible explanation uses the observation that the returns  $\rho_i$  are almost uncorrelated. Assuming  $b = P_n$  in (2), the price  $P_n$  after the  $n$ -th bid is given by a multiplicative process of the form  $P_n = (\varrho_n + 1)P_{n-1}$ . Since the returns  $\varrho_n$  are found to be uncorrelated,  $\ln(P_n) =$

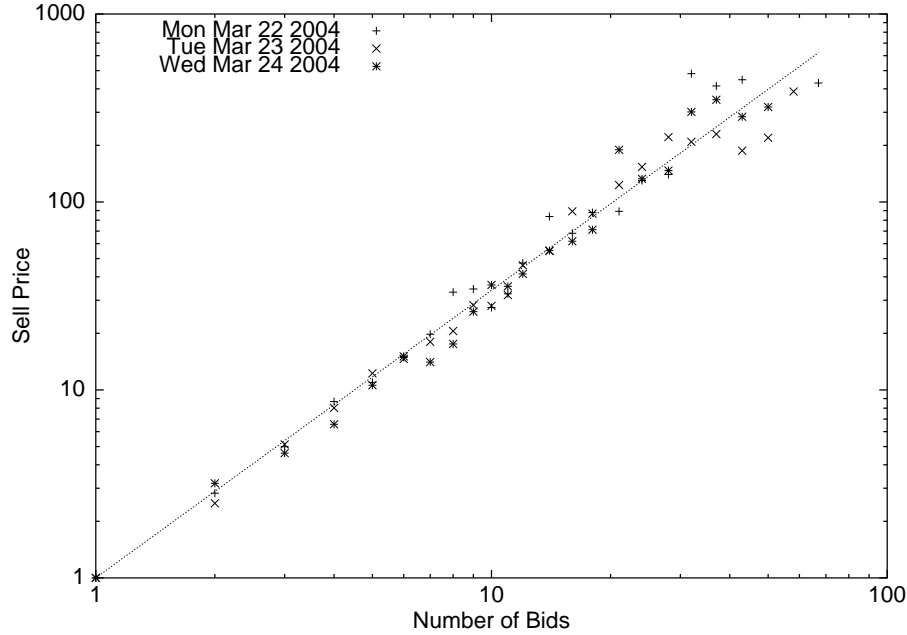


Fig. 2. Relation between final price  $\langle p_{\text{final}} \rangle$  (in Euro, for starting price 1 Euro) and number of bids  $n_{\text{bid}}$  for different days based on subsets of DB-1. The straight line is a power-law fit with exponent  $\alpha = 1.53$ .

$\sum_{i=1}^n \ln(\varrho_i + 1)$  converges to a Gaussian distribution for large  $n$ .

In the following we discuss a possible origin of systematic deviations from the average behaviour described above, namely fraud.

### 3.1. Shill bidding

It is not uncommon that certain agents try to manipulate the sell price of the items they offer. Therefore we distinguish two different types of bidding behaviour: 1) All bidders try to keep the price as low as possible; 2) At least one bidder tries to push the price higher. The second type of behaviour is known as shill bidding. Shill bidding (also known as "bid padding") is the deliberate use of secondary registrations, aliases, family members, friends, etc. to artificially drive up the price of an item. This is strictly forbidden by eBay<sup>10</sup>, but nevertheless happens quite frequently. Usually this sort of manipulation can be identified only after the auction has ended because the whole purpose of shill bidding is increasing the price without winning in the end! Shill bidding occurs mostly in auctions of products that do not have a well-defined value, e.g. market price.

We have tried to identify this sort of manipulation through the statistical properties. Indeed, successful shill bidding leads to auctions which show clear deviations from the observed simple statistical laws. To quantify this we have defined

a confidence interval based on the assumption that the price follows a log-normal distribution (see inset in Fig. 3). The variance  $\sigma^2$  of the Gaussian distribution of the logarithm of prices with a fixed number of bids characterizes the confidence interval which is indicated in Fig. 3 by the two straight lines with slopes 1.32 and 1.92, respectively.

We have performed two tests by a) investigating the statistical properties of auctions identified e.g. from discussion forums<sup>11</sup> as shill bidding, and b) by checking whether randomly chosen auctions outside the confidence interval of the price-bid relationship show indications of shill bidding. Both tests require an extensive amount of work, e.g. the investigation of the trading history of the seller over a long time or monitoring eBay discussion forums<sup>11</sup>. Fig. 3 shows a comparison between these shill-auctions and the average behaviour of all auctions (DB-1 and DB-2), clearly indicating systematic deviations.

For test a) we have chosen 9 auctions that clearly have been identified as manipulated by shill bidding, e.g. through information from discussion forums<sup>11</sup>. Only after that we have investigated the bidding history of these auctions in more detail. Fig. 3 shows that all of those, except for one, are clearly outside the confidence interval. For test b) 10 auctions outside the confidence interval have been chosen randomly. These have been checked thoroughly for indications of shill bidding. This also required investigating other auctions by the same seller etc. In this way we have found a clear indication for shill bidding in 7 of the 10 auctions.

The results of the tests indicate that it would be sufficient to check the suspicious auctions (i.e. those outside of the confidence interval) in more detail. This would reduce the complexity of detecting manipulations drastically. Furthermore these deviations usually can be observed even *before* the end of the auction. Monitoring the current price as function of the number of submitted bids up to the time of observation, one can check whether this price is inside or outside of the confidence interval in Fig. 3. In the latter case, a high current price is an indication that bidders should be cautious (although it is no proof of shill bidding).

### 3.2. Bid submission times

Second price auctions were originally introduced by Vickrey<sup>5</sup> to make everyones bidding strategy independent of strategies used by the other bidders. It could be shown that late bidding or bidding with a price less than one is willing to pay are not optimal strategies. Using game theory based models<sup>16,17,18,19</sup>, it seems hard to understand why bidders bid more than once. We analysed the collected data and found that bidders prefer to bid close to auction ending times. Fig. 4 shows the cumulative distribution of bid submission times as a function of the time remaining until the end of the auction. Two regimes with exponential behaviour can be observed related to the most common auction lengths of 7 and 10 days. Both parts are well described by  $P(\Delta t) \sim \exp(-\Delta t/T_0)$  with  $T_0 = 68.94$  h. Close to the end of the auctions so-called *sniping* leads to an algebraic distribution  $P(\Delta t) \sim (\Delta t)^{-\gamma}$

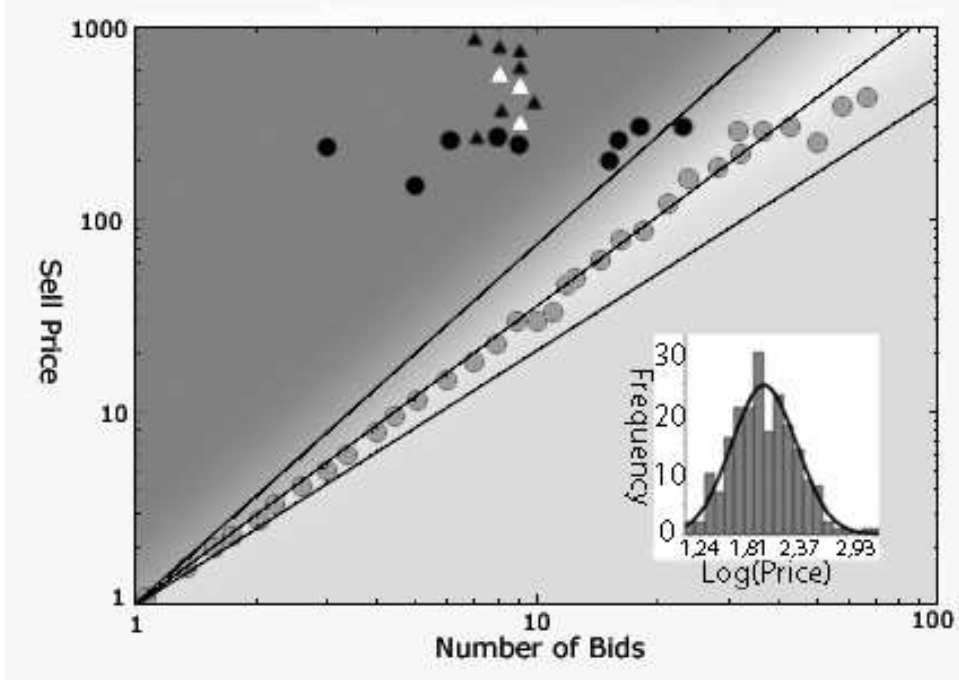


Fig. 3. Distribution of sell prices as function of the number of bids placed. Only auctions with a starting price of 1 Euro are considered. Gray dots correspond to the average sell price for fixed number of bids. The broken lines specify the confidence interval (one standard deviation) determined from a log-normal fit (see inset). This is shown for  $n_{\text{bid}} = 20$  in the inset. For other values of  $n_{\text{bid}}$  very similar results are obtained. Black dots indicate auctions identified as shill bidding using the criteria of Refs. 12-14 (test a). Triangles denote auctions that have been tested for possible shill bidding (test b). For black triangles strong indications for shill bidding have been found, whereas auctions corresponding to white triangles are unsuspecting. Light-gray and dark-gray colours denote regimes with high or low probability of shill bidding, respectively.

with  $\gamma = 1.1$ . Sniping is a special bidding strategy where the agent tries to submit a winning bid just before an auction closes in order to prevent other bidders from outbidding the sniper or driving the price higher. In fact, game-theoretic arguments show that sniping can be advantageous<sup>19</sup>, e.g. to avoid bid wars or auction fraud due to shill bidding.

Surprisingly, the distribution of bid submissions times is qualitatively very similar both data sets DB-1 and DB-2. Although the scales in the exponential part differ slightly ( $T_0 = 68.94$  h for DB-2 and  $T_0 = 90.09$  h for DB-1), the exponents of the algebraic part are almost identical. From earlier investigations<sup>16</sup> one could have expected a qualitative difference between the two data sets.

#### 4. Discussion

eBay online auctions can be considered as strongly correlated processes. Correlations are due to influences of standard market prices, supply and demand mechanisms

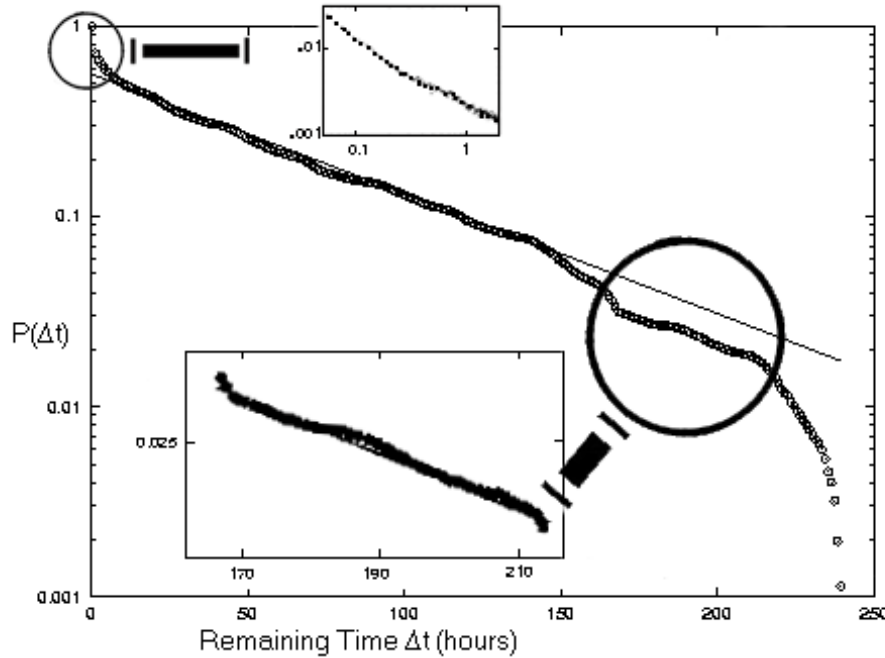


Fig. 4. Cumulative distribution of bid submission times as a function of the time remaining until the end of the auction (for DB-2). Two regimes with exponential behaviour can be observed related to the most common auction lengths of 7 and 10 days (i.e. 168 and 240 hours). Both parts are well described by  $P(\Delta t) \sim \exp(-\Delta t/T_0)$  with  $T_0 = 68.94$  h. Close to the end of the auctions sniping leads to an algebraic distribution  $P(\Delta t) \sim (\Delta t)^{-\gamma}$  with  $\gamma = 1.1$ . The data set DB-1 shows a very similar behaviour with  $T_0 = 90.09$  h and  $\gamma = 1.15$  (Ref. 6).

within eBay auctions and seller's feedback information available for any visitor of the site<sup>15</sup>. We have shown several emerging properties, which could be seen as an evidence of complexity. The generic statistic behavior of agents leads to a very simple relation between price and number of bids. As an application we have found that a kind of fraud known as shill bidding leads to significant deviations from the average behaviour.

Economic theories have mainly used game theoretic methods to describe auctions<sup>16,17,18,19</sup>. These are based on the assumption of *rational* agents using specific bidding strategies, like evaluation (placing one's true value, usually early in the auction), unmasking (bidding as long as someone else is the highest bidder, e.g. to determine the highest bid), or incremental bidding (placing minimum acceptable bids). Our results indicate that eBay can be regarded as a complex system consisting of a mixture of agents using different strategies. In future work we intend to determine the influence of the rationality of the agents (in the sense of game theory).

As mentioned in the introduction, to our knowledge so far no microscopic



stochastic model for the description of online auctions exists. Most approaches use game theoretical methods which focus on the effects of different strategies. Our empirical data indicate that a model needs to take into account that a mixture of agents with different strategies exists which is responsible for the observed behaviour. This can be seen most clearly in the distribution of bid submission times. In order to take into account the hard close of online auctions a naive approach would use dynamical rules which are itself time-dependent, which would be a challenge for theoretical studies of the model.

### Acknowledgments

We thank Axel Ockenfels and Maya Paczuski for helpful discussions.

### References

1. R.N. Mantegna and H.E. Stanley, *An Introduction to Econophysics* (Cambridge University Press, 2000).
2. J.P. Bouchaud and M. Potters, *Theory of Financial Risks* (Cambridge University Press, Cambridge, 2000)
3. I. Yang, H. Jeong, B. Kahng, and A.-L. Barabasi, Phys. Rev. E. **68**, 016102 (2003).
4. J. Reichardt, S. Bornholdt, physics/0503138
5. W. Vickrey, J. Finance **16**, 8 (1961)
6. A. Namazi, Dissertation, Universität zu Köln (2005)
7. M. I. Melnik and J. Alm, Jrl. Industrial Economics **50**(3), 337 (2002)
8. D. Lucking-Reiley, D. Bryan, N. Prasad, and D. Reeves: *Pennies from ebay: The determinants of price in online auctions*, Technical report, University of Arizona (2000)
9. W. Jank and G. Shmueli: *Dynamic Profiling of Online Auctions Using Curve Clustering*, unpublished (2005)
10. [pages.ebay.com/help/community/shillBidding.html](http://pages.ebay.com/help/community/shillBidding.html)
11. [answercenter.ebay.de](http://answercenter.ebay.de) (keywords: "shill", "pusher")
12. [www.darlingtontown.co.uk/hintsntips/Shill\\_bidding\\_on\\_eBay.shtml](http://www.darlingtontown.co.uk/hintsntips/Shill_bidding_on_eBay.shtml) (2004)
13. [www.ukauctionhelp.co.uk/shill.php](http://www.ukauctionhelp.co.uk/shill.php) (2004)
14. [www.basestealer.com/shilling.html](http://www.basestealer.com/shilling.html) (2004)
15. M. Melnik and J. Alm, Jrl. Indust. Econ. **50**(3), 337 (2002).
16. A. E. Roth and A. Ockenfels, Am. Econ. Rev. **92**, 1093 (2002)
17. A. E. Roth and A. Ockenfels, Games and Economic Behavior **92**, 1093 (2005)
18. R. T. Wilcox, Marketing Letters **11**, 363 (2000)
19. S. Barbaro and B. Bracht, [www.staff.uni-mainz.de/barbaro/node5.html](http://www.staff.uni-mainz.de/barbaro/node5.html)